ABSTRACT

This paper presents the evaluation of a mixed reality communication system for the home domain, called roomXT. The system uses a wall-sized display that is seamlessly integrated into a living lab, to create a ‘life-like’ video communication experience. In order to demonstrate the potential of this approach, we conducted a living lab study comparing the developed prototype with a desktop-based system. A special video communication application, which enables spatially separated users to have a joint dinner experience, served as a common basis for the different test conditions. Results of the study show that the overall concept of roomXT was well received by users of a wide age range and that the developed prototype system seems to be preferred to commercially available video communication solutions with respect to the tested quality dimensions.

Categories and Subject Descriptors
H.5.2 [Information interfaces and presentation]: User Interfaces—Prototyping, User-centered design, and Evaluation/methodology

General Terms
Design, Human Factors

Keywords
human-computer interaction, mixed-reality, large displays, co-dining, interactive media, family communication, co-presence, architecture

1. INTRODUCTION

Over the last decades, personal living situations changed profoundly in most industrialized countries. Social and demographic transformation processes led to changes in family structures, social networks and housing arrangements. The consequence of these developments are also reflected in the number of single households, which increased significantly in first world countries. Statistical data from the UK shows that there was an increase of single households in the working population aged 16 to 59 years from 5% in 1971 to 16% in 2002 [19]. There are reports on similar developments in other industrialized countries outside Europe, including the USA [9], Australia [3] and Japan [12].

However, the trend toward single households is not limited to the working population alone. Recent statistical data shows that the percentage of elderly people living alone increases dramatically with age. Approximately 44% of persons aged between 65 and 70 live in a single household [21] and it is expected that this number will further grow with the demographic change that is expected to take place in many western countries [17].

2. COMPUTER-MEDIATED INTERACTION

Traditional communication technologies, like telephone or e-mail, were originally designed to support goal-oriented information exchange between remote parties. A large portion of the commercial teleconferencing solutions as well as computer mediated cooperative work research focuses on working scenarios [8, 15]. Hence, it is not surprising that many of these technologies show short-comings when they are used in more intimate interaction scenarios [18]. Several projects have addressed this problem by designing dedicated...
communication devices for supporting awareness and informal communication among distributed family members and friends. These systems focus in particular on supporting social interactions among users in order to foster a feeling of connectedness within a group of closely related peers.

While most early systems, like, e.g., Sideshow [7], or Team Portal [2], were designed for supporting awareness and informal communication in distributed working groups, more recent approaches have a stronger focus on connecting users in technology-enhanced home environments. Well-known systems for supporting awareness in small intimate groups include FamilyPlanter [13], or Digital Family Portraits [14]. More recent systems following a cross-media approach include applications like Family Portals [10].

3. RESEARCH GOALS

Today, most of these systems are still built around dedicated communication devices, which come either in the form of shared stationary systems available in the users’ environment [20] or as mobile artifacts for personal usage. While this appears to be a promising design choice, empirical evidence suggests that many users prefer calm and unobtrusive technology that blends into the existing environment [6] instead of having visible communication devices within their living spaces [4]. In our work, we explicitly address this challenge and aim at providing users with a ‘natural’ communication channel for informal and spontaneous interactions.

With the roomXT prototype [11] we developed a shared communication space, which virtually extends the physical environment and thereby provides distant communications partners with a shared interaction space.

In order to demonstrate the validity of our approach, we compare the developed prototype system with a state-of-the-art communication application. This paper describes the user study that was conducted in order to explore the openness of future users to accept and use video communication systems for supporting and enhancing social interactions. More specifically, we investigated which quality dimensions of the video experience are responsible for forming the overall acceptance patterns. From a social point of view, we were also interested in whether the positive bias of a joint social event is reducing the technical character of the situation. As the nature of the technical installation might impact the acceptance, we compared a conventional desktop-based video application with different variations of our roomXT system.

4. EVALUATION CONCEPT

4.1 Remote Dining

As a first step towards evaluating roomXT, we set up an immersive video connection with a two-person dinner scenario. The dinner scenario has been chosen for several reasons. From our own experience, and from discussions with peers, we assume that people prefer dining in company. Most people that we talked to during pre-studies did see a strong connection between food consumption and social interaction. Some compensate for a missing dining partner by watching television, others already use video chats to accompany their dinner.

Another reason for choosing the dinner as our evaluation scenario was that everybody knows the situation of a real dinner with friends. The user can thus relate to this situation and compare our proposed system against the real situation. However, it is important to emphasize that we do not intend to substitute ‘real’ dining situations, but want to offer new opportunities for people maintaining a relationship over distance.

4.2 Communication Setup

Conceptually, we envision a symmetric system setup of two identically equipped rooms that offer the same experience for two dining partners (Figure 2, A). The cameras at each site should be placed in such a way that the table in the captured video shares the vanishing point of the local table, and that the user sits upright with the head at a defined position. At this one position, the user has the optical impression of the table extending into the screen. Once the user leaves this position, the vanishing point of the table in the video will differ from the vanishing point of the real table (see Figure 3, B), which disturbs the visual continuity. Depending on the position of the camera relative to the table and on the focal length of the camera, there will be areas on the table that are not captured by the camera (labeled "b" in Figure 2, A). The table area that is captured (labeled "v" in Figure 2, A) should include food and drinks as well as the area where the user places hands and arms. The perceived length of the table will be the length of the real table plus the visible part of the remote table. Figure 2A shows the camera attached to the wall, however, the ideal position of the cameras would be behind the walls [16]; in case of virtual site (Figure 2, C), the ideal camera position is at the head position of the user in the living lab. This explains the position of the camera in our setup at the remote site, (Figure 2, B) with the camera placed off the table. A placement like that allows setting a realistic focal length that matches the perception of the human eye. However, when aiming to achieve two identical setups, the more
realistic option in terms of technical feasibility is to keep the camera in front of the display and to use a wider focal length (see Figure 2, A).

4.3 Technical Setup

For the evaluation of the system, the communication scenario had to be slightly altered in order to increase experimental flexibility and to accommodate financial constraints. The communication infrastructure was installed in two separate locations within the same building.

The evaluation of the roomXT prototype system was conducted in a living lab environment (Figure 1). The lab contains a wall-sized display (4.8 x 2.4m, 3072 x 1536 pixels) that is realized using rear-projection. The frame size of the displayed content is varied depending on the evaluated condition (compare Figure 3, B and C). A table (90 x 90cm) is placed in front of the center of the display wall and audio and video are captured by a consumer level wide-angle web-camera with an integrated microphone located on the table (as shown in Figure 2, B left). In addition, a Microsoft Kinect sensor, situated at a distance of about 3 meters to the display wall, captures the area around the table. The Microsoft Kinect SDK is used to perform marker-less skeletal tracking. Smoothing the head position gives us jitter-free data that can be used to compute an OpenGL projection matrix.

A common office room functions as the remote site. Its purpose is to serve audio and video footage for the best possible video conferencing experience for the dining partner in the living lab (Figure 2, C illustrates the resulting experience). A 24-inch display is used to show the video stream from the living lab and a consumer wide-angle camera is placed above the display at the eye level of a sitting user. A greenscreen and a green table cloth are used to enable simple background segmentation in the recorded video.

A custom Java software is used for transmission of video data and 3D rendering. Low latency video streams at a resolution of 640 x 480 pixels are shown at 30 frames per second. In addition to the normal rendering of video images, the software also supports rendering of background segmented video images that are blended into a 3D scene.

Using the the Apple iChat application for audio transmission results in good quality and acoustic echo cancellation. Transmitting all data over our local network ensures low latency audio, so that a synchronization of audio and video data is not necessary. Overall, the quality of the video conferencing system seems to be superior to many consumer level solutions. In particular the low latency is relevant for natural human communication.

5. USER STUDY

To address our research goals, we perform an explorative study of six quality dimensions, which we assume to predict user acceptance. We expect that increasing the immersive-ness and using a deviceless system has a positive effect on user acceptance. Furthermore, we evaluate user characteristics and expect to receive lower acceptance scores from older users and users with a low experience in using technology.

5.1 Experimental Setting and Test Conditions

The following description of the experimental setting only applies to the living lab. The remote site was not part of the evaluation and the technical setup described earlier remained unchanged throughout the experiment.

Participants taking part in the study were asked to take a place at the table in front of the display wall in the living lab. Three courses where served during the experiment. Each course featured a different technical setting, which, in the following, we will refer to as conditions. The conditions differ in the way the video stream was presented; the sound setup was not changed during the study.

Figure 3: Experimental conditions. A: Desktop video application. B: Limited roomXT - no background segmentation, no head tracking. C: RoomXT - background segmentation and head tracking enabled.

In the desktop video application condition (Figure 3, A), the video stream of the dining partner was shown using a laptop placed on the table. It served as the baseline of the experiment with consumer hardware and software. This is the only condition that used a dedicated device for communication.

In the limited roomXT condition (Figure 3, B), the wall-sized display showed the unmodified and unmasked video stream at a size of 1.1m x 0.825m. The focal length and position of the remote camera were chosen in such a way that the perspective best matches the viewing position of a sitting person. From this position, the participant has a more or less perspective correct vision of the dining partner as well as of the remote area of the table where the food is located. The video stream of the remote table was positioned directly above the physical table, which aims at creating the impression of a spatial extension in which the physical table and the remote table are connected. However, moving the head away from the default position of a sitting person reduced the illusion of contiguity.

In the roomXT condition (Figure 3, C), we blended a masked video stream of the dining partner on top of a virtual 3D table (90 x 90cm). The goal was to create the impression that the user dining at the remote site is actually sitting on the virtual table. The masking hides anything but the dining partner and objects placed on the partner’s table. The two persons dining together thus shared a space, which is limited to the vicinity of the table. The video stream does not provide insights into the private living space. A neutral gray background color was rendered outside the mask. Additionally, we used head tracking to determine the correct perspective for rendering the virtual 3D table. As a result, the perspective of the real table seamlessly continued into the virtual table. The masked video stream was scaled and rendered at the position where the end of the table in the video coincides with the end of the virtual table. This results in that the dinner partner is constantly being rendered.

1See http://johnymlee.net/projects/wii/ for code examples of the calculation.
in a proportionally correct size. Because the table at the remote site was masked, too, we could simply render the video stream on top of the virtual table to create the impression that the unmasked objects were lying on the virtual table. It was thought that this setting should create the highest level of immersion.

5.2 Sample and User Characteristics

A total of 20 participants (10 pairs) aged between 22 and 50 years (60% female) volunteered to take part in the study. Most participants were recruited via the social networks of the authors. We explicitly looked for dinner partners who are highly familiar with each other (friends, partners, working colleagues). This was done to create a realistic atmosphere and to exclude any feeling of awkwardness between partners, which could have confounded the acceptance ratings of the technical conditions. Participants were invited to take part in a study about the use of video communication technology for shared dining. It was stressed that the environment is not aimed to replace face-to-face communication, but to provide additional functionality when communicating with remote friends and family members. Participants were highly educated (90% had a university degree).

In order to assess the participants’ expertise with technical devices, participants were asked whether they own a personal computer and mobile phone, how frequently they use them and how they rate the ease of using each device. Beyond technical expertise, we determined the subjective technical self-confidence (STC) by using a psychometric test developed by Beier [5], which revealed to be a sensitive variable in explaining technical performance [1] and acceptance towards novel technology [22]. Descriptive data showed that participants generally had high technical self-confidence (M = 67.7 out of 100 points, SD = 4.4). STC scores showed a significant correlation with gender (r = −.69; p < .05), with female participants (M = 66 points, SD = 5) having lower scores in contrast to male participants (M = 70 points, SD = 2.2).

In addition, we assessed the participants’ experience of using different social media technologies as well as their usage motives. With respect to the frequency of using personal computers and mobile phones, the sample revealed to be highly experienced in using technology. All participants (independently of age and gender) reported to use the devices on a daily basis, both for work and private purposes. The perceived ease of use was rated as “easy” or even “very easy”.

STC represents the perception of a person regarding the individual competence when using technical devices. Participants were given the short version of the test containing eight items (e.g., “Usually, I cope with technical problems successfully”), which had to be rated on a 5-point Likert scale, ranging from 1 (totally disagree) to 5 (totally agree). The maximum score that could be reached was 100 points. Descriptive data showed that participants generally had high STC (M = 67.7 points, SD = 4.4). STC scores showed a significant correlation with gender (r = −.69; p < .05), with female participants (M = 66 points, SD = 5) having lower scores in contrast to male participants (M = 70 points, SD = 2.2).

Participants were asked to indicate which and how often they would use social media to contact and communicate with friends and family members. 90% of participants reported to use social media technology regularly for private communication purposes, Facebook and Skype (with video function) being the most frequently used applications. In order to understand individual reasons for using social media, participants were asked to assess different usage motives on a 6-point Likert scale, ranging from 1 (totally disagree) to 6 (totally agree). The tested usage motives were determined in focus groups conducted prior to the present study. Figure 4 shows the results.

5.3 Design and Testing Procedure

The experiment was based on a 3 factorial design with repeated measurements. All participants were tested under all technical variations. Participants had a three-course meal (starter: Caprese; main: vegetarian lasagna; desert: vanilla flan with strawberries). The order of courses was constant, while the three conditions were varied across courses and participants. At first, participants filled out a questionnaire regarding personal data, technical expertise and familiarity with using social media technology. Then, the dinner started. After each condition, participants answered six questions each covering a different quality dimension. Participants evaluated if the condition (1) provided fun (“hedonism”) and (2) provoked a feeling of awkwardness, as well as (3) the novelty of the dinner experience, (4) the degree to which they found it difficult to manage using social media and to eat at the same time (“ease of handling multiple tasks”), (5) their wish to repeat such a “virtual dinner”, and (6) the degree to which they would like to use this environment regularly. Items were assessed on a 5-point Likert scale, ranging from 1 (totally disagree) to 5 (totally agree).

After the last course, participants were requested to rate the deliciousness of all three courses as well as the different technical conditions. The separate evaluation of the meal and the technical condition allowed us to estimate in how far the positive experience of enjoying a meal differs from the respective experience in the three different environments. The experiment lasted for approximately 30 minutes, depending on the individual communication behaviors and eating speeds.

5.4 Results

The results were analyzed by bivariate correlations (Spearman Rho) and multivariate analyses of variance (MANOVA) for repeated measurements as well as Friedman rank analyses for non-parametric data. The significance level was set to 5%; values within the less restrictive 10% level were
referred to as marginally significant. Post hoc tests were accomplished with the Bonferroni formula.

Descriptive outcomes of the addressed quality dimensions are depicted in Figure 5. In some of the quality dimensions, significant differences across the three technical installations of the dinner experience were revealed. The desktop video application showed the lowest novelty of dining experience ($\chi^2 = 3.9; p < .05$) compared to the limited roomXT and the roomXT condition (which did not differ significantly). Similarly, the wish to use the application again was smallest in the desktop video application, followed by both roomXT conditions ($\chi^2 = 3.7; p < .1$). Even though the experienced fun shows the same descriptive pattern (desktop video application last and both roomXT conditions superior), differences failed to reach statistical significance. The feeling of awkwardness did not differ across conditions, but was rated similarly. The same applies for the perceived ease of handling the multitasking and the wish to use the system on a regular basis (not only when family members or friends are living remote).

![Figure 5: Evaluation profiles of the three experimental conditions (5 = completely agree)](image)

5.4.1 Evaluation of Meal Quality vs. Conditions

In order to control whether the assessment of the meal is confounded with the evaluation of the different experimental conditions, participants rated both dimensions separately. For the evaluation of the meal, three grades were available (1=very good; 2=good; 3=satisfactory), for the evaluation of conditions we used a 6-point scale (1=very good, 6=very bad). On average, the meal was rated as “very good” ($M = 1.43$ points), with no differences between courses (starter: $M = 1.5$ ($SD = .52$); main: $M = 1.2$ ($SD = .42$); desert: $M = 1.4$ ($SD = .51$)). The experimental conditions were rated, on average, with 2.3 out of 6 points. Differences between conditions were significant ($\chi^2 = 3.11; p < .05$), revealing limited roomXT as the best condition ($M = 1.8$; $SD = .78$), followed by the desktop video condition ($M = 2.6$; $SD = .84$) and the roomXT condition ($M = 2.6; SD = 1.3$), which were both rated equally well. It is worth mentioning that participants were not as homogenous in their evaluation of the roomXT condition compared to the desktop video and limited roomXT condition (taken from the higher standard deviation). Apparently, participants are quite undecided whether head tracking and the immersiveness of the environment represent actual benefits for the tested dinner application. No significant correlations between evaluation dimensions were found, showing that the quality of the meal and the experimental conditions were evaluated differently, not confounding each other.

5.4.2 Effects of User Characteristics

A final analysis addressed the question whether user characteristics (technical expertise, experience of using social media technologies, age, gender) might impact the openness to use video communication technologies in domestic spaces. Spearman correlation analysis showed two statistically significant relations: one is the correlation between age and the openness of participants to use video conferencing in domestic spaces regularly ($r = -.71; p < .05$). The openness of participants for using digital environments for social purposes decreases with age. The other is the significant relation of subjective technical confidence (STC) and the evaluation of the desktop video condition ($r = .73x; p < .05$). People with a high level of STC accept the desktop video condition more strongly than persons with only a low level of STC. The fact that only the desktop video condition showed significant relations (and not the other conditions) implies that the limited roomXT as well as the roomXT condition is less sensitive to user diversity and, accordingly, more positively received by less technology-experienced people.

6. DISCUSSION AND CONCLUSION

Generally, the idea of combining video communication functionality and social activities within a domestic environment was evaluated as very positive by users of a wide age range. Users liked the joint dining experience and reported a high intention to use the system again. The results also show that the technical installation impacts the assessment of the system: desktop video applications were rated as less favorable compared to the large-scale limited roomXT video communication system. It is noteworthy that this not only applies for young people who are highly experienced with using technology, but also for middle-aged persons. Also, no gender differences regarding the openness to novel technological developments were identified. This finding contradicts results of numerous studies, which reported a lower acceptance of older persons towards technology [1] and the higher reluctance of women to use and handle novel technologies (e.g., [22]). Apparently, the life-like communication channel provided by the roomXT system was to a much lesser extent perceived as “technology-like” and therefore able to meet the needs and wants of a diverse user group.

In contrast to our expectation that the roomXT condition would increase the user experience, this was not the case. Beyond the higher ratings with respect to the novelty of the dining experience, which is superior in the roomXT condition, both the wall-sized limited roomXT and the roomXT condition showed a similarly high acceptance.

7. LIMITATIONS AND FUTURE WORK

Some final remarks deal with potential limitations of the experimental and technical approach. A first point refers to the selection of the sample. We are aware that the sample of highly educated and experienced participants represents
a rather untypical user group, which is not representative for the group of potential end users. It should therefore be kept in mind that a kind of “best case” scenario was under study. Future studies will have to validate if the high openness to use the system is still given in more representative groups. A second limitation considers the type of application domain. Eating and mealtimes are positively evaluated by nature, and therefore the whole application might have been benefited by this positive social bias. A third remark is directed to the fact that in our studies dinner partners were familiar with each other, relying on a developed relationship. However, there are many other contexts in which such systems could be used as, e.g., dating, which could impact the perceived usefulness, the experienced fun, and the feeling of awkwardness, which deserve further experimental work.

Furthermore, some technical limitations have to be considered. Creating a (seamless) wall-sized display is challenging. Our projector-based realization obviously has a lower quality in terms of resolution and colors than a standard laptop display, which is expected to have a negative effect on user preference. Using the Microsoft Kinect sensor for marker-less head tracking brings the advantage that our living lab setup remains free from a more complicated tracking setup, which is usually not found in a living room. One negative aspect of using marker-less tracking is the considerable amount of jitter in the detected skeleton positions. To receive a more robust head position, smoothing is necessary, which introduces a short delay. Future work should determine the effect of this delay on the perceived quality of the system. Finally, we assume that introducing more spacial 3D context, serving as reference frame, could increase the perceived quality of the system when head-tracking is enabled.

8. ACKNOWLEDGMENTS

We would like to thank all dinner volunteers. Many thanks also to Dominik Schlüter and Luisa Bremen for research assistance. This work was supported by the excellence initiative of the German federal and state government.

9. REFERENCES


